

CASE STUDY:

The Impact of AI Voice Moderation on the Call of Duty Player Experience

This case study outlines the findings and impact uncovered by Modulate and Activision through carefully designed analysis of ToxMod's impact on the Call of Duty community.



Executive Summary

Activision's Call of Duty franchise stands as a titan in the realm of first-person action games, captivating millions of players with its immersive gameplay and dynamic multiplayer experiences. Its legacy, built on a foundation of thrilling combat and strategic teamwork, has cemented it as a favorite among gamers worldwide.

In its online multiplayer modes which utilize voice chat for team communication and coordination, toxicity presents an ongoing challenge. In response, the Call of Duty team at Activision has implemented a series of advancements to the player experience, including the launch of a [new player reporting system](#), regular [progress reports](#) on anti-toxicity as well as ongoing anti-cheat efforts as part of its [RICOCHET Anti-Cheat initiative](#). Specific to combating toxicity in voice chats, the Call of Duty team collaborated with Modulate to implement its ToxMod technology.

ToxMod is specifically designed to address the unique challenges of moderating in-game voice communication. By leveraging machine learning tuned with real gaming data, ToxMod can tell the difference between competitive banter and genuine harassment.

This case study outlines the findings and impact uncovered by Modulate and Activision through carefully designed analysis of ToxMod's impact on the Call of Duty community.

While the primary focus of the analysis was to understand and improve player experience, working closely with the Call of Duty team and complementing additional related efforts, Modulate was able to analyze the impact the introduction of voice moderation was having on player engagement, and found sizable positive effects.



Within Call of Duty®: Modern Warfare® II and Call of Duty®: Warzone™
(North America only):



ToxMod identified **rates of toxicity and toxicity exposure in voice chats well above the rates** that existing player reports alone identified



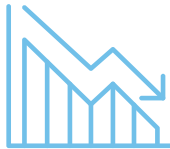
ToxMod found that only **~23% of player-generated reports contained actionable evidence** of a Code of Conduct violation ([Call of Duty Code of Conduct](#))



Thanks to the additional offenses identified by ToxMod, **Activision was better able to take action against offenders** – which in turn led to an increase in player engagement



Player churn was reduced when ToxMod was enabled



Toxicity exposure decreased by 25%

Within Call of Duty®: Modern Warfare® III (globally, excluding Asia),
ToxMod's impact included:

8%

Reduction in repeat offenders month-over-month

~50%

Reduction in toxicity exposure

2mil

Accounts for disruptive voice chat were acted upon by moderators in accordance with the Call of Duty Code of Conduct



Introduction

Multiplayer games have given people a unique way to connect with others through play. Voice chat-enabled games, in particular, have taken the player experience to a whole new level, adding a more human and more immersive layer to gameplay, fostering a greater sense of community across the globe.

One of these games that has left an undeniable mark on the multiplayer gaming space is Call of Duty published by Activision. Over the past two decades, Call of Duty has become one of the leading first-person action game in the world. The Guinness Book of World Records reconfirmed its standing as one of the top-selling video game following the launch of Call of Duty: Modern Warfare III in November 2023.

Games like Call of Duty are popular because they foster connection, competition, skill and fun. Prior to the official launch of ToxMod in Call of Duty, an ADL report found that 77% of adult video game players had experienced some form of severe harassment — and Call of Duty is most definitely not immune. And with a fanbase of this size, moderating that toxicity presents unique challenges.

An ADL report found that 77% of adult video game players had experienced some form of severe harassment.

Modulate's ToxMod aims to reduce players' exposure to harmful content through proactive, ML-driven voice moderation, thereby contributing toward improving player engagement and retention. ToxMod allows moderation teams to deploy advanced, complementary efforts on often-unactionable, player-generated reports that lead toward a more proactive moderation strategy — a pivotal move in the ongoing battle against in-game toxicity.

What's the Problem?

Toxic behavior, ranging from derogatory remarks to harassment, not only tarnishes individual gameplay experiences, but also can erode the sense of camaraderie and respect that underpins healthy gaming communities.

The impact of such behavior extends beyond momentary discomfort; it can lead to players taking a step away from the game for a few hours, days, or even quitting altogether (also known as player churn) and diminished community engagement. As Activision continued to fulfill its initiatives to support Call of Duty's player community, the teams at Activision and Modulate developed a hypothesis:

Shifting toward proactive voice moderation via ToxMod would materially improve player experience, while materially reducing toxicity exposure rates.

Next, it was time to put that hypothesis to the test by integrating ToxMod.

ToxMod's Integration into Call of Duty

Recognizing the limitations of traditional moderation methods and the unique challenges presented by real-time voice communication, the decision to adopt ToxMod was driven by Activision's commitment to maintaining a positive and inclusive gaming environment for the Call of Duty community.

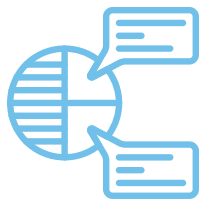
This partnership ensured that ToxMod's advanced voice moderation capabilities were seamlessly woven into the existing game infrastructure, with minimal impact on game performance and user experience.

Key considerations included:

- Careful tuning to adhere to Activision's [Call of Duty Code of Conduct](#).
- Preserving the competitive and fast-paced spirit of gameplay.
- Compatibility with the game's diverse gameplay modes.
- Adherence to privacy standards and privacy laws.
- Scalability to accommodate the massive Call of Duty player base.
- Maintaining lowest possible latency for toxicity detection.

How ToxMod Works Within Call of Duty

ToxMod operates within Call of Duty through a sophisticated, multi-stage process designed to proactively identify and prioritize toxic voice chat interactions for Activision's human moderator team. ToxMod is also designed to respect player privacy. To that end, ToxMod is designed to recognize speech, but ToxMod does not engage in speaker identification, and does not create a biometric voiceprint of any user. This process can be broken down into three phases:



Triage

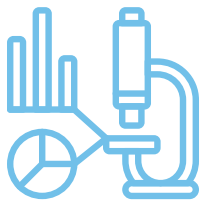
In the first stage, ToxMod analyzes voice communications in real-time, looking for toxic speech as defined by Call of Duty's Code of Conduct. This initial filtering allows ToxMod to determine which conversations warrant greater attention and is crucial for efficiently identifying conversations that warrant closer examination, ensuring that the system remains focused on the most likely problematic interactions.



Analyze

Interactions flagged in the triage stage then undergo a deeper analysis to understand context and intention. It evaluates nuances: slang, tone of voice, cultural references, and the conversation between players. By doing so, ToxMod can distinguish between competitive banter, which is a natural part of the gaming experience, and genuinely harmful content. With this information, ToxMod can better uncover key context of a voice interaction so a moderator can determine the next course of action. ToxMod focuses on phrases or slurs which are unequivocally bad and undergoes the following types of analysis:

- **Emotion and speech:** recognizes emotions, including anger, which can help differentiate between the banter typical (and welcome!) in Call of Duty and genuine hurt or aggression.
- **Sentiment analysis:** analyzes the full utterance in context of the broader conversation (both before and after the utterance itself) to better understand the intent and sentiment with which it was spoken.



Escalate

After ToxMod prioritizes and analyzes a voice chat interaction that is very likely a violation of Call of Duty's Code of Conduct, the issue is escalated to Activision for review. Rather than funneling all voice chat interactions to moderators, this tiered approach ensures that potential false positives are removed from the moderation flow. Moderator actions can range from issuing warnings to temporary or permanent communication bans, depending on the severity of the offense.

Initial Analysis Results

ToxMod's impact was initially assessed within North America for English-speaking Modern Warfare II and Call of Duty: Warzone players. This initial analysis allowed Activision teams to gather initial insights into the scale and type of behavior happening in voice chats and to fine-tune ToxMod's detection specifically for the Call of Duty player base. Activision tested manual moderation actioning based on ToxMod's detection on a treatment group and maintained a control group where ToxMod would still detect likely Code of Conduct violations, but no moderator action would be taken.

Toxicity Exposure

In the control group, ToxMod's data showed at least 25% of the Modern Warfare II player base was exposed to severe gender/sexual harassment (~90% of detected offenses) and racial/cultural harassment (~10% of detected offenses). Where was toxicity coming from?

Among all voice chat infractions in the treatment group, ToxMod data shows that about 50% of infractions were from first-time offenders. Analysis showed that of the total warnings issued to players for first-time detected offenses, the vast majority were issued to players who were already active in Call of Duty— that is to say, players who are already regularly playing Call of Duty titles. Only ~10% of first-time offense warnings were issued to new players or players returning to Call of Duty after some time:

2.1%

of first-time offense warnings were given to **new players of Call of Duty**

4.7%

of first-time offense warnings were given to lapsed players who **returned to Call of Duty after 21-59 days absence**

1.7%

of first-time offense warnings were given to players who **returned to Call of Duty after 60+ days absence**

During this analysis period, Activision adopted a three-tiered enforcement flow, with a 48-hour cooldown before players could be escalated into the next enforcement tier:

Tier 1

Player is **sent a warning** about their voice chat behavior violating the Call of Duty Code of Conduct

Tier 2

Player is **muted for 3 days** and notified

Tier 3

Player is **muted for 14 days** and notified

19% of toxicity exposure was due to players violating the Code of Conduct while in a cooldown period following a moderator warning. 22% of toxicity exposure was due to players violating the Code of Conduct after a moderator penalty had been lifted. Within the repeat offenses, 13% of those offenses occurred after tier-1 warning, 7% after tier-2 shadow mute for 3 days and notified, 2% after a tier-3 shadow mute for 14 days and notified.

In periodic tests comparing exposure to toxicity in the treatment group and the control group, ToxMod was consistently found to reduce toxicity exposure between 25-33%.

Reactive Player Reports

Modulate and Activision also looked at the efficacy of reactive moderation in the form of player-generated reports. Data showed that reactive moderation approaches like player-generated reports addressed only a small fraction of the violations. For example:



On average, approximately 79% of players violating the Code of Conduct and escalated by ToxMod each day have no associated player reports – these offenders might not ever have been found without ToxMod's proactive detection



Approximately 50% of player reports submitted had no associated audio from reported players in voice chat 24 hours before the report was made



Of the reports with associated audio, only an estimated 50% of them will contain a Code of Conduct violation – this suggests that only about one quarter of player reports contained actionable evidence of toxicity in voice chat

Player Engagement

Modulate and Activision also analyzed the impact of proactive voice moderation on player engagement. Proactive moderator actioning against Code of Conduct violations in the treatment group boosted the overall number of active players in the treatment group. Comparing the treatment group to the control group in Modern Warfare II, the treatment group saw:

+3.9%
more new players

+2.4%
more players who were
previously inactive
for 21-59 days

+2.8%
more players who were
previously inactive
for 60+ days

Notably, the longer our moderation efforts went on, the larger the positive impact and more players remaining active in the game. Modulate and Activision teams compared the total number of active players in the treatment group to the control group after 3 days, 7 days and 21 days from the start of the testing period and found the treatment group saw:

+6.3%
more active players on **day 3**

+21.2%
more active players on **day 7**

+27.9%
more active players on **day 21**

Global Launch Results

Using ToxMod data, Activision was able to report on the [results of proactive moderation](#) in Call of Duty: Modern Warfare III following the game's launch in November 2023 in all regions across the globe except Asia. Here are some key findings:

A more potent reduction to toxic voice chat exposure

Call of Duty saw a ~50% reduction in players exposed to severe instances of disruptive voice chat since Modern Warfare III's launch. This decrease highlights the progress being made by Activision and Modulate since the trial period. Not only does it show that players are having a much better time online, it also speaks to improvements in overall player engagement.

A decrease in repeat offenders

ToxMod's ability to identify and help moderators take action against toxic players led to an 8% reduction in repeat offenders month over month, contributing to a healthier community dynamic. This 8% reduction in repeat offenders in Modern Warfare III shows that as ToxMod continues to run, more and more players recognize the ways in which their actions violate the Code of Conduct, and learn to adapt their behavior to something less exclusionary or offensive.

An increase in moderator enforcement of the Call of Duty Code of Conduct

More than 2 million accounts have seen in-game enforcement for disruptive voice chat, based on the Call of Duty Code of Conduct between August and November 2023. Of the severe toxicity that ToxMod flagged, only 1 in 5 were also reported by players, meaning that **ToxMod enabled Activision to catch, and ultimately put a stop to, 5x more harmful content** without putting any extra burden on Call of Duty players themselves to submit a report.

Conclusion

The integration of ToxMod into the most popular video game franchise in the world represents a significant step in Activision's ongoing efforts to reduce toxicity in Call of Duty titles. Beyond Call of Duty, Activision's strong stance against toxicity demonstrates what is possible for other game franchises across the globe, redefining in-game communication standards and setting a new benchmark for proactive moderation in the multiplayer gaming industry. By prioritizing real-time intervention and fostering a culture of respect and inclusivity, Call of Duty is not only enhancing the gaming experience for its players but also leading by example in the broader gaming industry.